

PRICING AMERICAN OPTIONS VIA MULTI-LEVEL APPROXIMATION METHODS ¹

Denis Belomestny, Fabian Dickmann and Tigran Nagapetyan
Duisburg-Essen University and Fraunhofer ITWM

In this article we propose a novel approach to reduce the computational complexity of various approximation methods for pricing discrete time American options. Given a sequence of continuation values estimates corresponding to different levels of spatial approximation and time discretization, we propose a multi-level low biased estimate for the price of an American option. It turns out that the resulting complexity gain can be unexpectedly high and can even reach the order ε^{-2} with ε denoting the desired precision. The performance of the proposed multilevel algorithm is illustrated by a numerical example of pricing Bermudan max-call options.

1 INTRODUCTION

Pricing an American option usually reduces to solving an optimal stopping problem which can be efficiently solved in low dimensions via dynamic programming algorithm. However, many problems arising in practice (see e.g. Glasserman (2004)) have high dimensions, and these applications have motivated the development of Monte Carlo methods for pricing American option. Pricing American style derivatives via Monte Carlo is a challenging task because it requires a backwards dynamic programming algorithm that seems to be incompatible with the forward structure of Monte Carlo methods. In recent years much research was focused on the development of fast methods to compute approximations to the optimal exercise policy. Eminent examples include the functional optimization approach of Andersen (2000), the mesh method of Broadie and Glasserman (1997), the regression-based approaches of Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (1999), Egloff (2005) and Belomestny (2011). The complexity of the fast approximations algorithms depends on the desired precision ε in a quite nonlinear way which in turn is determined by some fine properties of the underlying exercise boundary and the continuation values (see, e.g., Belomestny (2011)). In some situations (e.g. in the case of the stochastic mesh method or local regression) this complexity is of order ε^{-4} which is rather high. One way to reduce the complexity of the fast approximation methods is to use various variance reduction methods. However, the latter methods are often ad hoc and, more importantly, do

¹This research was partially supported by the Deutsche Forschungsgemeinschaft through the SPP 1324 “Mathematical methods for extracting quantifiable information from complex systems” and by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

not lead to provably reduced asymptotic complexity. In this paper we propose a generic approach which is able to reduce the order of the asymptotic complexity and which is applicable to various fast approximation methods, such as global regression, local regression or stochastic mesh method. The main idea of the method is inspired by the pathbreaking work of Giles (2008) which introduced a multilevel idea into stochastics. As similar to the recent work of Belomestny et al (2012), we consider not only levels corresponding to different discretization steps but also levels related to different degrees of approximation of the continuation values. For example, in the case of the Longstaff-Schwartz algorithm the latter degree is basically governed by the number of basis functions and in the case of the mesh method by the number of “training paths” used to approximate the continuation values. The new multi-level approach is able to significantly reduce the complexity of the fast approximation methods leading in some cases to the complexity gain of the order ε^{-2} . The paper is organised as follows. In Section 2 the pricing problem is formulated, the main assumptions are introduced and illustrated. In Section 3 the complexity analysis of a generic approximation algorithm is carried out. The main multi-level Monte Carlo algorithm is introduced in Section 4 where also its complexity is studied. In Section 5 we numerically test our approach for the problem of pricing Bermudan max-call options via mesh method. The proofs are collected in Section 6.

2 MAIN SETUP

An American option grants the holder the right to select the time at which to exercise the option, and in this differs from a European option which may be exercised only at a fixed date. A general class of American option pricing problems can be formulated through an \mathbb{R}^d Markov process $\{X_t, 0 \leq t \leq T\}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$. It is assumed that the process (X_t) is adapted to $(\mathcal{F}_t)_{0 \leq t \leq T}$ in the sense that each X_t is \mathcal{F}_t measurable. Recall that each \mathcal{F}_t is a σ -algebra of subsets of Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \leq t$. We restrict attention to options admitting a finite set of exercise opportunities $0 = t_0 < t_1 < t_2 < \dots < t_{\mathcal{J}} = T$, sometimes called Bermudan options. Then

$$Z_j := X_{t_j}, \quad j = 0, \dots, \mathcal{J},$$

is a Markov chain. If exercised at time t_j , $j = 1, \dots, \mathcal{J}$, the option pays $g_j(Z_j)$, for some known functions $g_0, g_1, \dots, g_{\mathcal{J}}$ mapping \mathbb{R}^d into $[0, \infty)$. Let \mathcal{T}_j denote the set of stopping times taking values in $\{j, j+1, \dots, \mathcal{J}\}$. A standard result in the theory of contingent claims states that the equilibrium price $V_j(z)$ of the American option at time t_j in state z given that the option was not exercised prior to t_j is its value under an optimal exercise policy:

$$V_j^*(z) = \sup_{\tau \in \mathcal{T}_j} E[g_{\tau}(Z_{\tau}) | Z_j = z], \quad z \in \mathbb{R}^d.$$

A common feature of all fast approximation algorithms is that they can deliver estimates $C_{k,0}(z), \dots, C_{k,\mathcal{J}-1}(z)$ for the so called continuation values:

$$C_j^*(z) := E[V_{j+1}^*(Z_{j+1})|Z_j = z], \quad j = 0, \dots, \mathcal{J} - 1, \quad (2.1)$$

based on the set of trajectories $(Z_0^{(i)}, \dots, Z_{\mathcal{J}}^{(i)})$, $i = 1, \dots, k$, all starting from one point, i.e., $Z_0^{(1)} = \dots = Z_0^{(k)}$. In the case of the so-called regression methods, the estimates for the continuation values are obtained via the recursion (dynamic programming principle):

$$\begin{aligned} C_{\mathcal{J}}^*(z) &= 0, \\ C_j^*(z) &= E[\max(g_{j+1}(Z_{j+1}), C_{j+1}^*(Z_{j+1}))|Z_j = z] \end{aligned}$$

combined with Monte Carlo: at $(\mathcal{J} - j)$ th step one estimates the expectation:

$$E[\max(g_{j+1}(Z_{j+1}), C_{k,j+1}(Z_{j+1}))|Z_j = z] \quad (2.2)$$

by regression (global or local) based on the sample

$$(Z_j^{(i)}, C_{k,j+1}(Z_{j+1}^{(i)})), \quad i = 1, \dots, k,$$

where $C_{k,j+1}(z)$ is an estimate for $C_{j+1}^*(z)$ obtained in the previous step. Another way to approximate the continuation values $C_0(z), \dots, C_{\mathcal{J}-1}(z)$ is to maximize a Monte Carlo estimate of the expectation $E[g_{\tau_\theta}(Z_\tau)|Z_j = z]$ based on k paths of Z over a vector of parameters $\theta = (\theta_1, \dots, \theta_{\mathcal{J}}) \in \Theta^{\mathcal{J}}$ with

$$\tau_\theta = \min\{0 \leq l \leq j : \phi(Z_l, \theta_l) \leq g_l(Z_l)\},$$

where ϕ is a predefined function on $\mathbb{R}^d \times \Theta$. In this way one gets an estimate $\theta_k = (\theta_{k,1}, \dots, \theta_{k,\mathcal{J}})$ and defines $C_{k,j}(z) = \phi_j(z, \theta_{k,j})$.

Let us now consider a generic family of the estimates $C_{k,0}(z), \dots, C_{k,\mathcal{J}-1}(z)$ with a natural number k determining the quality of the estimates as well as their complexity. In particular we make the following assumptions.

(AP) For any $k \in \mathbb{N}$ the estimates $C_{k,0}(z), \dots, C_{k,\mathcal{J}-1}(z)$ are defined on some filtered probability space $(\Omega^k, \mathcal{F}^k, \mathbb{P}^k)$ which is independent of $(\Omega, \mathcal{F}, \mathbb{P})$.

(AC) For any $j = 1, \dots, \mathcal{J}$, and for any fixed $z \in \mathbb{R}^d$, the estimate $C_{k,j}(z)$ has numerical complexity of order k^α for some $\alpha > 0$.

(AQ) There is a sequence of positive real numbers γ_k with $\gamma_k \rightarrow 0$, $k \rightarrow \infty$ such that

$$\mathbb{P}^k \left(\sup_z |C_{k,j}(z) - C_j^*(z)| > \eta \sqrt{\gamma_k} \right) < B_1 e^{-B_2 \eta}, \quad \eta > 0$$

for some constants $B_1 > 0$ and $B_2 > 0$.

Let us now illustrate the above assumptions for three well known approximation methods.

Example 1 (Global regression). Fix a vector of real-valued functions $\psi = (\psi_1, \dots, \psi_L)$ on \mathbb{R}^d . Let $\alpha_j^k = (\alpha_{j,1}^k, \dots, \alpha_{j,L}^k)$ be a solution of the following least squares optimization problem:

$$\arg \inf_{\alpha \in \mathbb{R}^k} \sum_{i=1}^k \left[\zeta_{j+1,k}(Z_{j+1}^{(i)}) - \alpha_1 \psi_1(Z_j^{(i)}) - \dots - \alpha_L \psi_L(Z_j^{(i)}) \right]^2 \quad (2.3)$$

with $\zeta_{j+1,k}(z) = \max \{g_{j+1}(z), C_{k,j+1}(z)\}$. Define an approximation for C_j^* via

$$C_{k,j}(z) = \alpha_{j,1}^k \psi_1(z) + \dots + \alpha_{j,L}^k \psi_L(z), \quad z \in \mathbb{R}^d.$$

It is clear that all estimates $C_{k,j}$ are well defined on the cartesian product of k copies of (Ω, \mathcal{F}, P) . The complexity $\text{comp}(\alpha_j^k)$ of computing α_j^k is of order $k \cdot L^2 + \text{comp}(\alpha_{j+1}^k)$, since each α_j^k is of the form $\alpha_j^k = B^{-1}b$ with

$$B_{p,q} = \frac{1}{k} \sum_{i=1}^k \psi_p(Z_j^{(i)}) \psi_q(Z_j^{(i)})$$

and

$$b_p = \frac{1}{k} \sum_{i=1}^k \psi_p(Z_j^{(i)}) \zeta_{k,j+1}(Z_{j+1}^{(i)}),$$

$p, q \in \{1, \dots, L\}$. Hence $\text{comp}(\alpha_j^k) \sim (\mathcal{J} - j) \cdot k \cdot L^2$. Furthermore, it can be shown that the estimates $C_{k,0}(z), \dots, C_{k,\mathcal{J}-1}(z)$ satisfy the assumption (AQ) under some regularity conditions (see, e.g., Egloff (2005)), provided L increases with k in a logarithmic rate.

Example 2 (Local regression). Local polynomial regression estimates can be defined as follows. Fix some j such that $0 \leq j < \mathcal{J}$ and suppose that we want to compute the expectation in (2.2):

$$\mathbb{E}[\zeta_{j+1,k}(Z_{j+1}) | Z_j = z], \quad z \in \mathbb{R}^d$$

with $\zeta_{j+1,k}(z) = \max \{g_{j+1}(z), C_{k,j+1}(z)\}$. For some $\delta > 0$, $z \in \mathbb{R}^d$, an integer $l \geq 0$ and a function $K : \mathbb{R}^d \rightarrow \mathbb{R}_+$, denote by $q_{z,k}$ a polynomial on \mathbb{R}^d of degree l (i.e. the maximal order of the multi-index is less than or equal to l) which minimizes

$$\sum_{i=1}^k \left[\zeta_{j+1,k}(Z_{j+1}^{(i)}) - q(Z_j^{(i)} - z) \right]^2 K \left(\frac{Z_j^{(i)} - z}{\delta} \right) \quad (2.4)$$

over the set of all polynomials q of degree l . The local polynomial estimator of order l for $C_j^*(z)$ is then defined as $C_{k,j}(z) = q_{z,k}(0)$ if $q_{z,k}$ is the unique

minimizer of (2.4) and $C_{k,j}(z) = 0$ otherwise. Hence, for any $j = 0, \dots, \mathscr{J} - 1$, the complexity of computing the value $C_{k,j}(z)$ is of order k as $k \rightarrow \infty$. The value δ is called the bandwidth and the function K is called the kernel function. In Belomestny (2011) it is shown that the local polynomial estimates $C_{k,0}(z), \dots, C_{k,\mathscr{J}-1}(z)$ of degree l satisfy the assumption (AQ) under some regularity conditions, provided $\delta = k^{-1/(2l+d)}$.

Example 3 (Mesh Method). In the mesh method of Broadie and Glasserman (2004) the continuation value C_j^* at a point z is approximated via

$$C_{k,j}(z) = \frac{1}{k} \sum_{i=1}^k \zeta_{k,j+1}(Z_{j+1}^{(i)}) \cdot w_{ij}(z),$$

where $\zeta_{k,j+1}(z) = \max\{g_{j+1}(z), C_{k,j+1}(z)\}$ and

$$w_{ij}(z) = \frac{p_j(z, Z_{j+1}^{(i)})}{\frac{1}{k} \sum_{i=1}^k p_j(Z_j^{(i)}, Z_{j+1}^{(i)})},$$

where $p_j(x, y)$ is the transition density from $Z_j = x$ to $Z_{j+1} = y$. Hence, for any $j = 0, \dots, \mathscr{J} - 1$, the complexity of computing $C_{k,j}(z)$ is of order k , provided the transition density $p_j(x, y)$ is analytically known.

Based on the estimates $C_{k,0}(z), \dots, C_{k,\mathscr{J}-1}(z)$ one can construct a lower bound (low biased estimate) for V_0^* using the (generally suboptimal) stopping rule:

$$\tau_k = \min\{0 \leq j \leq \mathscr{J} : C_{k,j}(Z_j) \leq g_j(Z_j)\}$$

with $C_{k,\mathscr{J}} \equiv 0$ by definition. Fix two natural numbers N and K , and simulate N trajectories of the process Z . A low-biased estimate for V_0^* can be then defined via

$$V_0^{N,K} = \frac{1}{N} \sum_{r=1}^N g_{\tau_K^{(r)}}(Z_{\tau_K^{(r)}}^{(r)}),$$

where

$$\tau_K^{(r)} = \inf\{0 \leq j \leq \mathscr{J} : g_j(Z_j^{(r)}) > C_{K,j}(Z_j^{(r)})\}.$$

DISCRETIZATION Usually the process X and hence Z can not be simulated exactly and the so-called discretization schemes have to be used. For the sake of concreteness consider a d -dimensional diffusion process whose dynamics is given by

$$X_t = x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad (2.5)$$

where W is a standard d' -dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P_x)$ satisfying the usual conditions. The mappings b and σ are Lipschitz continuous in space and locally bounded in time, so that (2.5) has a unique strong solution. We approximate the diffusion (2.5) on the grid $0 = t_0 < t_1 < \dots < t_{\mathcal{J}}$ via Euler scheme with a discretization step $h > 0$ and discretization points ih , $i \in \mathbb{N}$. For $t \geq 0$ we define $\phi(t) = ih$ for $ih \leq t < (i+1)h$ and introduce

$$\tilde{X}_{h,t} = x + \int_0^t b(\phi(s), \tilde{X}_{h,\phi(s)}) ds + \int_0^t \sigma(\phi(s), \tilde{X}_{h,\phi(s)}) dW_s. \quad (2.6)$$

Put $\tilde{Z}_{h,j} = \tilde{X}_{h,t_j}$, $j = 0, \dots, \mathcal{J}$. Let now $h_k, k \in \mathbb{N}$, be a sequence of discretization steps tending to 0. Define

$$\tilde{V}_0^{N,K} = \frac{1}{N} \sum_{r=1}^N g_{\tilde{\tau}_K^{(r)}}(\tilde{Z}_{h_K, \tilde{\tau}_K^{(r)}}^{(r)}) \quad (2.7)$$

with

$$\tilde{\tau}_K^{(r)} = \inf \left\{ 0 \leq j \leq \mathcal{J} : g_j(\tilde{Z}_{h_K, j}^{(r)}) > C_{K,j}(\tilde{Z}_{h_K, j}^{(r)}) \right\}$$

Although the estimate (2.7) is not any longer low-biased due to a discretization error, it still can be viewed as a good approximation for $\tilde{V}_0^K := E[g_{\tau_K}(Z_{\tau_K})]$, provided h_K is small enough. In the next section we analyze the numerical complexity of the estimate $\tilde{V}_0^{N,K}$.

3 COMPLEXITY ANALYSIS OF $\tilde{V}_0^{N,K}$

In order to carry out the complexity analysis of the estimate (2.7) we need the so-called “margin” or boundary assumption.

(AM) There exist constants $A > 0$, $\delta_0 > 0$ and $\alpha > 0$ such that

$$P\left(|C_j^*(Z_j) - g_j(Z_j)| \leq \delta\right) \leq A\delta^\alpha$$

for all $j = 0, \dots, \mathcal{J}$, and all $\delta < \delta_0$.

Assumption (AM) provides a useful characterization of the behavior of the continuation values $\{C_j^*\}$ and payoffs $\{g_j\}$ near the exercise boundary $\partial \mathcal{E}$ with

$$\mathcal{E} = \left\{ (j, x) : g_j(x) \geq C_j^*(x) \right\}.$$

In the situation when all functions $C_j^* - g_j$, $j = 0, \dots, \mathcal{J} - 1$, are smooth and have non-vanishing derivatives in the vicinity of the exercise boundary, we have $\alpha = 1$. Other values of α are possible as well (see Belomestny (2011)). While the variance of the estimate $\tilde{V}_0^{N,K}$ is given by

$$\text{Var}[\tilde{V}_0^{N,K}] = \text{Var}[g_{\tilde{\tau}_K}(\tilde{Z}_{h_K, \tilde{\tau}_K})]/N,$$

its bias is analyzed in the following theorem.

Theorem 4. Suppose that (AM) and (AQ) hold, and all functions g_j are uniformly bounded and Lipschitz continuous, i.e.,

$$|g_j(x)| \leq G, \quad |g_j(x) - g_j(y)| \leq \mathcal{L}_g \|x - y\|, \quad x, y \in \mathbb{R}^d.$$

for some constants $G > 0$ and $\mathcal{L}_g > 0$. Moreover assume that all continuation functions C_j^* are uniformly Lipschitz, i.e.,

$$|C_j^*(x) - C_j^*(y)| \leq \mathcal{L}_C \|x - y\|, \quad x, y \in \mathbb{R}^d,$$

for $j = 0, \dots, \mathcal{J}$, and $k \in \mathbb{N}$. If

$$\limsup_{\delta \rightarrow 0} \mathbb{E} \left[\sup_{l \leq j \leq \mathcal{J}} |Z_j|^p \mid |C_l^*(Z_l) - g_j(Z_l)| \leq \delta \right] < \infty,$$

for $l = 0, \dots, \mathcal{J}$ and $p > \alpha$, then it holds

$$\left| V_0^* - \mathbb{E}[\tilde{V}_0^{N,K}] \right| \lesssim \gamma_K^{(1+\alpha)/2} + \left(\gamma_K \log^2 \frac{1}{\gamma_K} \vee h_K \log^2 \frac{1}{h_K} \right)^{\alpha/2} + h_K^{1/2}, \quad K \rightarrow \infty.$$

The next theorem gives an upper estimate for the complexity of $V_0^{N,K}$.

Theorem 5. Let assumptions of Theorem 4 hold and

$$\gamma_k = k^{-\mu}, \quad k \in \mathbb{N}$$

for some $\mu > 0$. Then under the choice $h_k = k^{-\beta}$ with

$$\beta = \begin{cases} \mu, & 0 < \alpha \leq 1, \\ \mu\alpha, & 1 < \alpha \end{cases}$$

and for any $\delta > 0$ the complexity of the estimate (2.7) given

$$\mathbb{E} [\tilde{V}_0^{N,K} - V_0^*]^2 \leq \varepsilon^2,$$

is bounded from above by the value $\mathcal{C}_{N,K}(\varepsilon)$ with

$$\begin{cases} \varepsilon^{-4-\frac{2\alpha}{\mu\alpha}}, & \alpha \geq 1, \\ \varepsilon^{-2-\frac{2}{\alpha}-\frac{2\alpha}{\mu\alpha}}, & 0 \leq \alpha < 1, \end{cases} \lesssim \mathcal{C}_{N,K}(\varepsilon) \lesssim \begin{cases} \varepsilon^{-4-\frac{2\alpha}{\mu\alpha}-\delta(\alpha+\mu\alpha)}, & \alpha \geq 1, \\ \varepsilon^{-2-\frac{2}{\alpha}-\frac{2\alpha}{\mu\alpha}-\delta(\alpha+\mu)}, & 0 \leq \alpha < 1, \end{cases}$$

provided $\varepsilon^{-2/(\mu\alpha)} \lesssim K \lesssim \varepsilon^{-2/(\mu\alpha)-\delta}$.

DISCUSSION Theorem 5 implies that the complexity of the estimate $\tilde{V}_0^{N,K}$ can be rather high and can even reach the order ε^{-q} for arbitrary large $q > 0$. In the next section we introduce a multilevel approach which is able to reduce the asymptotic complexity order.

4 MULTILEVEL APPROACH

Fix some natural number L and let $\mathbf{k} = (k_1, \dots, k_L)$ and $\mathbf{n} = (n_1, \dots, n_L)$ be two sequences of natural numbers. Define

$$\tilde{V}_0^{\mathbf{n}, \mathbf{k}} = \frac{1}{n_0} \sum_{r=1}^{n_0} g_{\tilde{\tau}_{k_0}^{(r)}} \left(\tilde{Z}_{h_{k_0}, \tilde{\tau}_{k_0}^{(r)}}^{(r)} \right) + \sum_{l=1}^L \frac{1}{n_l} \sum_{r=1}^{n_l} \left[g_{\tilde{\tau}_{k_l}^{(r)}} \left(\tilde{Z}_{h_{k_l}, \tilde{\tau}_{k_l}^{(r)}}^{(r)} \right) - g_{\tilde{\tau}_{k_{l-1}}^{(r)}} \left(\tilde{Z}_{h_{k_{l-1}}, \tilde{\tau}_{k_{l-1}}^{(r)}}^{(r)} \right) \right]$$

with

$$\tilde{\tau}_k^{(r)} = \inf \left\{ 0 \leq j \leq \mathcal{J} : g_j(\tilde{Z}_{h_k, j}^{(r)}) > C_{k,j}(\tilde{Z}_{h_k, j}^{(r)}) \right\}, \quad k \in \mathbb{N}.$$

The following theorem gives the bias and the variance of the estimate $\tilde{V}_0^{\mathbf{n}, \mathbf{k}}$.

Theorem 6. *Let (AQ) and (AM) hold with some $\alpha > 0$, then the estimate $\tilde{V}_0^{\mathbf{n}, \mathbf{k}}$ has the bias of the order*

$$\gamma_K^{(1+\alpha)/2} + \left(\gamma_K \log^2 \frac{1}{\gamma_K} \vee h_K \log^2 \frac{1}{h_K} \right)^{\alpha/2} + h_K^{1/2}$$

and the variance of the order

$$\frac{\text{Var}[g(X_{h_{k_0}, \tau_{k_0}})]}{n_0} + \sum_{l=1}^L \frac{1}{n_l} \left(\left(\gamma_{k_{l-1}} \log^2 \frac{1}{\gamma_{k_{l-1}}} \vee h_{k_{l-1}} \log^2 \frac{1}{h_{k_{l-1}}} \right)^{\alpha/2} + h_{k_{l-1}} \right).$$

Furthermore, under assumption (AC) the complexity of $\tilde{V}_0^{\mathbf{n}, \mathbf{k}}$ is bounded from above by a multiple of

$$\sum_{l=0}^L k_l^x n_l h_{k_l}^{-1}.$$

Finally, the complexity of $\tilde{V}_0^{\mathbf{n}, \mathbf{k}}$ is given by the following theorem.

Theorem 7. *Let assumptions of Theorem 4 hold and let*

$$\gamma_{k_l} = k_l^{-\mu}, \quad k_l \in \mathbb{N}$$

for some $\mu > 0$. Then under the choice $h_{k_l} = k_l^{-\beta}$ and $k_l = k_0 \kappa^l$, $l = 0, 1, \dots, L$, with

$$\beta = \begin{cases} \mu, & 0 < \alpha \leq 1 \wedge 1 < \alpha \leq 2 \vee \frac{2x}{\mu\alpha} \leq 1, \\ \mu\alpha, & 1 < \alpha \vee \frac{2x}{\mu\alpha} \geq 1, \\ \frac{\mu\alpha}{2}, & 2 < \alpha \vee \frac{2x}{\mu\alpha} \leq 1, \end{cases}$$

$$L = \begin{cases} \left\lceil \left[\frac{2}{\mu\alpha} \log_{\kappa} \varepsilon^{-1} + \frac{3}{\mu} \log_{\kappa} \left(\log_{\kappa} \varepsilon^{-\frac{2}{\alpha}} \right) \right] \right\rceil, & 0 < \alpha \leq 1, \\ \left\lceil \frac{2}{\beta} \log_{\kappa} \varepsilon^{-1} \right\rceil, & 1 < \alpha, \end{cases}$$

the complexity of the estimate (2.7), given that

$$\mathbb{E} \left[\tilde{V}_0^{\mathbf{n}, \mathbf{k}} - V_0^* \right]^2 \leq \varepsilon^2,$$

is bounded, up to a constant, from above by $\mathcal{C}_{\mathbf{n}, \mathbf{k}}(\varepsilon)$ from Figure 4.

Figure 4.1: Complexity for the multilevel approach.

$$\mathcal{C}_{\mathbf{n},\mathbf{k}}(\varepsilon) = \begin{cases} \begin{cases} \varepsilon^{-2-\frac{4x}{\mu\alpha}} \log_{\kappa}^{1+\alpha} \varepsilon^{-1}, & \frac{2x}{\mu\alpha} \leq 1, \\ \varepsilon^{-3-\frac{2x}{\mu\alpha}} \log_{\kappa}^{1+\frac{3}{\mu}} \varepsilon^{-1}, & \frac{2x}{\mu\alpha} \geq 1, \end{cases} & \alpha > 2, \\ \begin{cases} \varepsilon^{-4+\alpha-\frac{2x}{\mu}} \log_{\kappa} \varepsilon^{-1}, & \frac{2x}{\mu\alpha} \leq 1, \\ \varepsilon^{-3-\frac{2x}{\mu\alpha}} \log_{\kappa} \varepsilon^{-1}, & \frac{2x}{\mu\alpha} \geq 1, \end{cases} & 1 < \alpha \leq 2, \\ \varepsilon^{-1-\frac{2}{\alpha}-\frac{2x}{\mu\alpha}} \log_{\kappa}^{\min\{1+\frac{3}{\mu}, 4\}} \varepsilon^{-1}, & 0 < \alpha \leq 1. \end{cases}$$

DISCUSSION Let us compare the complexities of the estimates $\tilde{V}_0^{N,K}$ and $\tilde{V}_0^{\mathbf{n},\mathbf{k}}$. To this end we compute the ratio function

$$\mathcal{R}(\varepsilon) = \frac{\mathcal{C}_{\mathbf{n},\mathbf{k}}(\varepsilon)}{\mathcal{C}_{N,K}(\varepsilon)}.$$

As can be easily seen, the largest complexity gain with $\mathcal{R}(\varepsilon) \approx \varepsilon^2$ up to a logarithmic factor can be, for example, attained in the situation $\frac{2x}{\mu\alpha} \approx 0$, $\alpha > 1$, which in turn takes place if $\alpha = \infty$ and $\mu > 0$, since for all known approximation algorithms $x \leq 1$. An example of pricing problems where the assumption (AM) holds with an arbitrary large α can be found in Belomestny (2011).

5 NUMERICAL EXAMPLE: BERMUDAN MAX CALLS ON MULTIPLE ASSETS

Suppose that the price of the underlying asset $X = (X^1, \dots, X^d)$ follows a Geometric Brownian motion (GBM) under the risk-neutral measure, i.e.,

$$dX_t^i = (r - \delta)X_t^i dt + \sigma X_t^i dB_t^i, \quad (5.1)$$

where r is the risk-free interest rate, δ the dividend rate, σ the volatility, and $B_t = (B_t^1, \dots, B_t^d)$ is a vector of d independent standard Brownian motions. At any time $t \in \{t_0, \dots, t_{\mathcal{J}}\}$ the holder of the option may exercise it and receive the payoff

$$h(X_t) = e^{-rt} (\max(X_t^1, \dots, X_t^d) - \kappa)^+.$$

We consider a benchmark example (see, e.g. Glasserman (2004), p. 462) when $d = 2$, $t_j = jT/\mathcal{J}$, $j = 0, \dots, \mathcal{J}$, with $T = 3$ and $\mathcal{J} = 9$.

5.1 MESH METHOD

Fix some natural numbers k_0 and L , and define a sequence of natural numbers via

$$k_l = k_0 \times 10^l, \quad l = 0, \dots, L.$$

For each $l = 1, \dots, L$, simulate independently two k_l “training” paths of the process Z using the exact formula

$$Z_j^{(i)} = Z_{j-1}^{(i)} \exp \left(\left[r - \delta - \frac{1}{2} \sigma^2 \right] (t_j - t_{j-1}) + \sigma \sqrt{(t_j - t_{j-1})} \cdot \xi_j^i \right),$$

where ξ_j^i , $i = 1, \dots, k_l$, are i. i. d. standard normal random variables. The corresponding transition density is given by

$$p_j(x, y) = \prod_{i=1}^d p_j(x_i, y_i), \quad x = (x_1, \dots, x_d), \quad y = (y_1, \dots, y_d),$$

where

$$p_j(x_i, y_i) = \frac{x_i}{y_i \sigma \sqrt{2\pi(t_j - t_{j-1})}} \exp \left(\frac{- \left(\log \left(\frac{y_i}{x_i} \right) - \left(r - \delta - \frac{1}{2} \sigma^2 \right) (t_j - t_{j-1}) \right)^2}{2\sigma^2(t_j - t_{j-1})} \right).$$

Using the above paths we construct the sequence of estimates

$$C_{k_l,0}(x), \dots, C_{k_l,\mathscr{J}}(x), \quad l = 1, \dots, L,$$

as described in Example 3. Next fix a sequence of natural numbers $n_0 < n_1 < \dots < n_L$ and consider the estimate

$$\begin{aligned} V_0^{\mathbf{n},\mathbf{k}} &= \frac{1}{n_0} \sum_{r=1}^{n_0} g_{\tau_0^{(r)}} \left(Z_{\tau_0^{(r)}}^{(r)} \right) \\ &\quad + \sum_{l=1}^L \frac{1}{n_l} \sum_{r=1}^{n_l} \left[g_{\tau_l^{(r)}} \left(Z_{\tau_l^{(r)}}^{(r)} \right) - g_{\tau_{l-1}^{(r)}} \left(Z_{\tau_{l-1}^{(r)}}^{(r)} \right) \right] \end{aligned}$$

with

$$\tau_l^{(r)} = \inf \left\{ 0 \leq j \leq \mathscr{J} : g_j(Z_j^{(r)}) \geq C_{k_l,j}(Z_j^{(r)}) \right\},$$

where $(Z_0^{(r)}, \dots, Z_{\mathscr{J}}^{(r)})$, $r = 1, \dots, n_l$, is a set of n_l paths of the process Z . Furthermore one can use one and the same set of k_l “training paths” to estimate both $C_{k_l,j}$ and $C_{k_{l-1},j}$, $l = 1, \dots, L$. This would reduce both the variance and the complexity of $V_0^{\mathbf{n},\mathbf{k}}$. The complexity of the estimate $\tilde{V}_0^{\mathbf{n},\mathbf{k}}$ is proportional to

$$\mathscr{C}_L(\mathbf{n}, \mathbf{k}) = n_0 k_0 + \sum_{l=1}^L (n_l k_l + n_l k_{l-1})$$

and its variance is given by

$$\mathscr{V}_L^2(\mathbf{n}, \mathbf{k}) = \frac{\sigma_0^2}{n_0} + \sum_{l=1}^L \frac{\sigma_l^2}{n_l},$$

Table 1: The estimated level variances for different values of k_L together with $\sigma_L^* = \mathcal{V}_L(\mathbf{n}^*, \mathbf{k})$.

$L \backslash k_L$	2.5×10^3	5×10^3	10×10^3	20×10^3
0	$\sigma_0 = 12.058$ $\sigma_0^* = 0.0539$	$\sigma_0 = 12.084$ $\sigma_0^* = 0.0764$	$\sigma_0 = 11.982$ $\sigma_0^* = 0.1071$	$\sigma_0 = 11.965$ $\sigma_0^* = 0.1513$
1	$\sigma_0 = 11.841$ $\sigma_1 = 5.838$ $\sigma_1^* = 0.0441$	$\sigma_0 = 11.963$ $\sigma_1 = 5.342$ $\sigma_1^* = 0.0593$	$\sigma_0 = 12.023$ $\sigma_1 = 4.622$ $\sigma_1^* = 0.0773$	$\sigma_0 = 12.040$ $\sigma_1 = 3.998$ $\sigma_1^* = 0.1011$
2	$\sigma_0 = 10.533$ $\sigma_1 = 7.015$ $\sigma_2 = 5.882$ $\sigma_2^* = 0.0427$	$\sigma_0 = 10.890$ $\sigma_1 = 6.690$ $\sigma_2 = 5.274$ $\sigma_2^* = 0.0559$	$\sigma_0 = 11.416$ $\sigma_1 = 6.291$ $\sigma_2 = 4.672$ $\sigma_2^* = 0.0727$	$\sigma_0 = 11.739$ $\sigma_1 = 5.828$ $\sigma_2 = 3.984$ $\sigma_2^* = 0.0921$
$V_0^{\mathbf{n}^*, \mathbf{k}}$	7.9799	8.0245	8.0464	8.0678

with

$$\sigma_0^2 = \text{Var} [g_{\tau_0}(Z_{\tau_0})], \quad \sigma_l^2 = \text{Var} [g_{\tau_l}(Z_{\tau_l}) - g_{\tau_{l-1}}(Z_{\tau_{l-1}})], \quad l = 1, \dots, L.$$

First we simulate k_l “training” paths, $n = 10000$ “testing” paths and use 100 repetitions of “training” and “testing” steps to estimate the level variance σ_l^2 for all $l = 0, \dots, L$. The estimated level variances are presented in Table 1. Next for any $L = 0, 1, 2$, we fix $\mathcal{C}_0 = 125 \times 10^6$ and numerically solve the optimisation problem:

$$\mathbf{n}^* = \arg \min_{\mathbf{n} \in \mathcal{N}} \mathcal{V}_L(\mathbf{n}, \mathbf{k}), \quad \mathcal{N} = \{n \in \mathbb{N}^L : \mathcal{C}_L(\mathbf{n}, \mathbf{k}) \leq \mathcal{C}_0\}.$$

In this way we find the optimal vector \mathbf{n}^* leading to the smallest variance of the estimate $V_0^{\mathbf{n}^*, \mathbf{k}}$ under the budget constraints. The results are presented in Table 1 with $\sigma_L^* = \mathcal{V}_L(\mathbf{n}^*, \mathbf{k})$, $L = 1, 2, 3$. The values of $V_0^{\mathbf{n}^*, \mathbf{k}}$ are also given. In Figure 5.1 the corresponding ratios $(\sigma_L^*)^2/(\sigma_0^*)^2$, $L = 0, 1, 2$, are shown.

5.2 IMPORTANCE SAMPLING

One can significantly improve the efficiency of the multilevel approach by applying the importance sampling technique. Let us fix some $l > 0$ and look at the distribution of the r. v.

$$\Delta_l = h_{\tau_l}(X_{\tau_l}) - h_{\tau_{l-1}}(X_{\tau_{l-1}}).$$

As can be seen from the Figure 5.2, Δ_l vanishes for 80%-90% of the “testing”

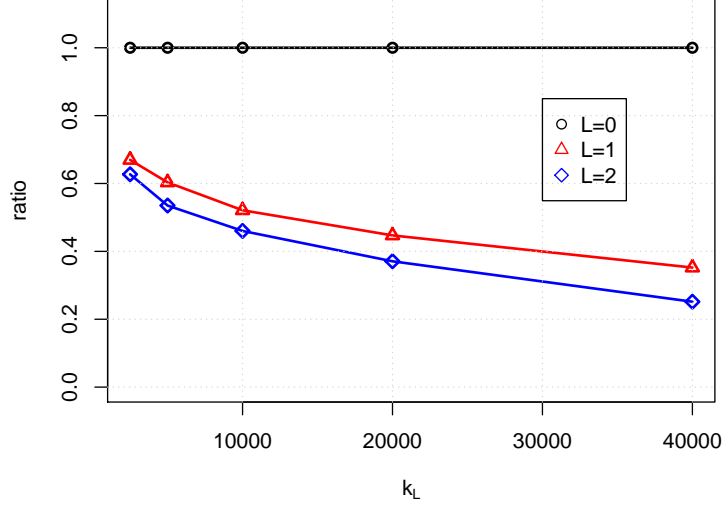


Figure 5.1: Variance reduction effect of the ML approach: the ratio of variances $(\sigma_L^*)^2/(\sigma_0^*)^2$ for $L = 0, 1, 2$.

paths in our example and this motivates the application of importance sampling technique. First we change the measure from P to Q via

$$\frac{dQ}{dP}(\omega) = \begin{cases} 1/P(\mathcal{J}_l), & \omega \in \mathcal{J}_l, \\ 0, & \omega \in \Omega \setminus \mathcal{J}_l, \end{cases}$$

where

$$\{h_{\tau_l}(X_{\tau_l}) \neq h_{\tau_{l-1}}(X_{\tau_{l-1}})\} \subset \mathcal{J}_l := \{\tau_l \neq \tau_{l-1}\} \subset \Omega.$$

Now, for a set of testing paths $X^{(r)}$, $r = 1, \dots, n_l$, generated under Q the unbiased Monte-Carlo estimator for $E_P[\Delta_l]$ is given by

$$\bar{\Delta}_l^n = \frac{1}{n_l} \sum_{r=1}^{n_l} P(\mathcal{J}_l) \left\{ h_{\tau_l}^{(r)}(X_{\tau_l}^{(r)}) - h_{\tau_{l-1}}^{(r)}(X_{\tau_{l-1}}^{(r)}) \right\} \quad (5.2)$$

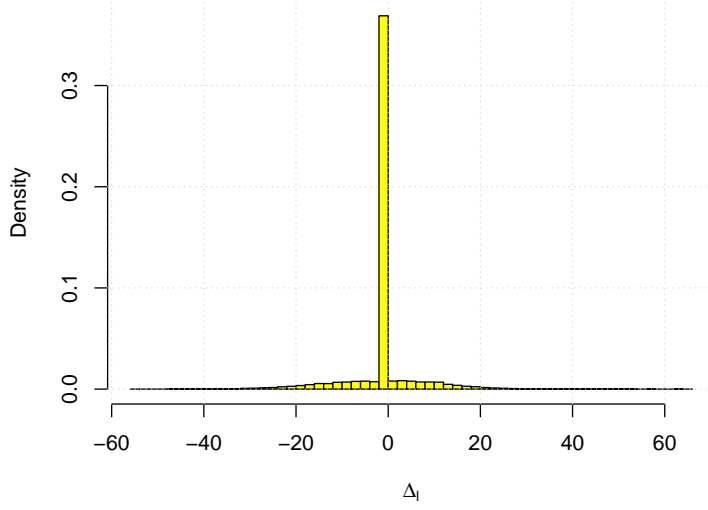


Figure 5.2: Histogram of the r.v. Δ_l based on 10000 realisations. An atom in 0 is clearly visible.

Moreover, it holds

$$\begin{aligned}
 \text{Var}_Q[\Delta_l] &= E_Q[\Delta_l^2] - E_Q^2[\Delta_l] \\
 &= E_P \left[\Delta_l^2 \frac{dQ}{dP} \right] - E_P^2 \left[\Delta_l \frac{dQ}{dP} \right] \\
 &= \frac{1}{P(\mathcal{J}_l)} \cdot E_P[\Delta_l^2] - \left(\frac{1}{P(\mathcal{J}_l)} \cdot E_P[\Delta_l] \right)^2 \\
 &= \frac{1}{P(\mathcal{J}_l)} \left(\text{Var}_P[\Delta_l] + E_P^2[\Delta_l] \right) - \left(\frac{1}{P(\mathcal{J}_l)} E_P[\Delta_l] \right)^2 \\
 &= \frac{1}{P(\mathcal{J}_l)} \text{Var}_P[\Delta_l] + E_P^2[\Delta_l] \underbrace{\left(\frac{1}{P(\mathcal{J}_l)} - \frac{1}{P(\mathcal{J}_l)^2} \right)}_{\leq 0}
 \end{aligned} \tag{5.3}$$

and as a consequence

$$\text{Var}_Q[\Delta_l] \leq \frac{1}{P(\mathcal{J}_l)} \text{Var}_P[\Delta_l].$$

In fact, the last inequality is quite tight, as $E_P^2[\Delta_l]$ in (5.3) will be negligible. As a result we have

$$\text{Var}_Q[\bar{\Delta}_l^n] \leq P(\mathcal{J}_l) \cdot \frac{\sigma_l^2}{n_l}, \quad (5.4)$$

meaning that importance sampling reduces the variance by a factor of at least $P(\mathcal{J}_l)$. Now we calculate again σ_L^* by optimisation of n_1, \dots, n_L where we used $\frac{\sigma_l^2}{P(\mathcal{J}_l)}$ instead of σ_l^2 for $l = 1, \dots, L$. The new ratios $(\sigma_L^*)^2/(\sigma_0^*)^2$ are shown in Figure 5.3. As \mathcal{J}_l is not known explicitly, sampling from Q is not directly

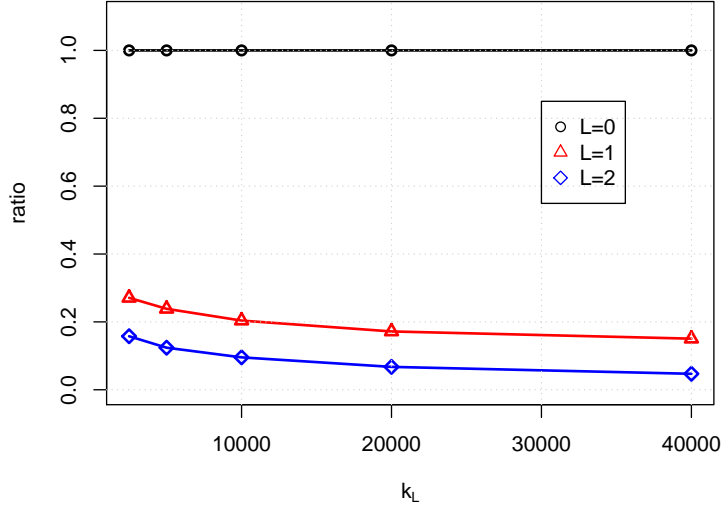


Figure 5.3: Variance reduction effect of the ML approach enhanced with importance sampling: the ratio of variances $(\sigma_L^*)^2/(\sigma_0^*)^2$ for $L = 0, 1, 2$.

possible. We apply the following strategy to obtain testing paths $Y_{\cdot}^{(r)}$, $r = 1, \dots, n_l$ that have approximately the distribution Q . We fix a natural number $1 < R \leq n_l$ and start to generate trajectories $Y_{\cdot}^{(1)}, Y_{\cdot}^{(2)}, \dots$ under P . If a path $Y_{\cdot}^{(q)}$ leads to different stopping times (i.e. it enters the symmetric difference of the exercise regions $\mathcal{E}(\tau_l) \Delta \mathcal{E}(\tau_{l-1})$ at timestep s), we will generate the next R paths $Y_{\cdot}^{(q+1)}, \dots, Y_{\cdot}^{(q+R)}$ starting from $Y_s^{(q)}$ at time s . This ensures that those paths will also lead to different stopping times. Afterwards we proceed again by generating paths $Y_{\cdot}^{(q+R+1)}, \dots$ under P and so on. In summary, the algorithm is described below.

1. Set $k := 0$, $\bar{k} := 0$, $r := 0$ and repeat the following steps while $r \leq n_l$:
 - (a) Set $r := r + 1$ and generate $Y_{\cdot}^{(r)} \sim P$.

(b) Calculate $\tau_l^{(r)} := \tau_l(Y^{(r)})$ and $\tau_{l-1}^{(r)} := \tau_{l-1}(Y^{(r)})$ and set

$$s := \min(\tau_l^{(r)}, \tau_{l-1}^{(r)})$$

(c) If $\tau_l^{(r)} = \tau_{l-1}^{(r)}$, set $\bar{k} := \bar{k} + 1$ and goto (a). Otherwise set $k := k + 1$ and repeat step (d) R times:

(d) Set $r := r + 1$ and generate trajectory $Y^{(r)}$ via

$$Y_t^{(r)} := \begin{cases} Y_t^{(r-1)} & 0 \leq t \leq s, \\ \tilde{Y}_t^{(r)} & s < t \leq \mathcal{J}, \end{cases} \quad (5.5)$$

where $\tilde{Y}_t^{(r)}$ is a trajectory starting from time s at $Y_s^{(r-1)}$ generated under $P(\cdot \mid Y_0^{(r-1)}, \dots, Y_s^{(r-1)})$. Calculate $\tau_l^{(r)} := \tau_l(Y^{(r)})$ and $\tau_{l-1}^{(r)} := \tau_{l-1}(Y^{(r)})$.

2. Define an estimator $\tilde{\Delta}_l^n$ for $E_P[\Delta^l]$ by

$$\tilde{\Delta}_l^n := \frac{1}{k + \bar{k}} \frac{1}{R + 1} \left\{ h_{\tau_l^{(r)}}(Y_{\tau_l^{(r)}}^{(r)}) - h_{\tau_{l-1}^{(r)}}(Y_{\tau_{l-1}^{(r)}}^{(r)}) \right\}.$$

3. Estimate $P(\mathcal{A}_l)$ by

$$\hat{p} = \frac{k}{k + \bar{k}}.$$

6 PROOFS

6.1 PROOF OF THEOREM 4

A family of stopping times $(\tau_j)_{j=0, \dots, \mathcal{J}}$ w.r.t. the filtration $(\mathcal{F}_j)_{j=0, \dots, \mathcal{J}}$ is called consistent if

$$j \leq \tau_j \leq \mathcal{J}, \quad \tau_{\mathcal{J}} = \mathcal{J}$$

and

$$\tau_j > j \implies \tau_j = \tau_{j+1}.$$

Lemma 8. Let $(Y_j)_{j=0, \dots, \mathcal{J}}$ be a process adapted to the filtration $(\mathcal{F}_j)_{j=0, \dots, \mathcal{J}}$ and let (τ_j^1) and (τ_j^2) be two consistent families of stopping times. Then

$$E^{\mathcal{F}_j} [Y_{\tau_j^1} - Y_{\tau_j^2}] = E^{\mathcal{F}_j} \left\{ \sum_{l=j}^{\mathcal{J}-1} \left(Y_l - E^{\mathcal{F}_l} [Y_{\tau_{l+1}^1}] \right) \left(1_{\{\tau_l^1=l, \tau_l^2>l\}} - 1_{\{\tau_l^1>l, \tau_l^2=l\}} \right) 1_{\{\tau_l^2>l\}} \right\}$$

for any $j = 0, \dots, \mathcal{J} - 1$.

Proof. We have

$$\begin{aligned}
Y_{\tau_j^1} - Y_{\tau_j^2} &= \left[Y_j - Y_{\tau_j^2} \right] \mathbf{1}_{\{\tau_j^1=j, \tau_j^2>j\}} + \left[Y_{\tau_j^1} - Y_j \right] \mathbf{1}_{\{\tau_j^1>j, \tau_j^2=j\}} \\
&\quad + \left[Y_{\tau_j^1} - Y_{\tau_j^2} \right] \mathbf{1}_{\{\tau_j^1>j, \tau_j^2>j\}} \\
&= \left[Y_j - Y_{\tau_{j+1}^1} \right] \mathbf{1}_{\{\tau_j^1=j, \tau_j^2>j\}} + \left[Y_{\tau_{j+1}^1} - Y_j \right] \mathbf{1}_{\{\tau_j^1>j, \tau_j^2=j\}} \\
&\quad + \left[Y_{\tau_{j+1}^1} - Y_{\tau_{j+1}^2} \right] \mathbf{1}_{\{\tau_j^1=j, \tau_j^2>j\}} + \left[Y_{\tau_{j+1}^1} - Y_{\tau_{j+1}^2} \right] \mathbf{1}_{\{\tau_j^1>j, \tau_j^2>j\}}.
\end{aligned}$$

Therefore it holds for $\Delta_j = \mathbb{E}^{\mathcal{F}_j} \left[Y_{\tau_j^1} - Y_{\tau_j^2} \right]$

$$\Delta_j = \left[Y_j - \mathbb{E}^{\mathcal{F}_j} \left[Y_{\tau_{j+1}^1} \right] \right] \left(\mathbf{1}_{\{\tau_j^1=j, \tau_j^2>j\}} - \mathbf{1}_{\{\tau_j^1>j, \tau_j^2=j\}} \right) + \mathbb{E}^{\mathcal{F}_j} \left\{ \Delta_{j+1} \mathbf{1}_{\{\tau_j^2>j\}} \right\}$$

with $\Delta_{\mathcal{J}} = 0$ and

$$\Delta_j = \mathbb{E}^{\mathcal{F}_j} \left\{ \sum_{l=j}^{\mathcal{J}-1} \left(Y_l - \mathbb{E}^{\mathcal{F}_l} \left[Y_{\tau_{l+1}^1} \right] \right) \left(\mathbf{1}_{\{\tau_l^1=l, \tau_l^2>l\}} - \mathbf{1}_{\{\tau_l^1>l, \tau_l^2=l\}} \right) \mathbf{1}_{\{\tau_l^2>l\}} \right\}.$$

□

Introduce

$$\bar{V}_0^{N,K} = \frac{1}{N} \sum_{r=1}^N g_{\tau_K^{(r)}} \left(\tilde{Z}_{h_K, \tau_K^{(r)}}^{(r)} \right)$$

and

$$V_0^{N,K} = \frac{1}{N} \sum_{r=1}^N g_{\tau_K^{(r)}} \left(Z_{\tau_K^{(r)}}^{(r)} \right),$$

$\bar{V}_0^{N,K}$ is an estimate computed using “exact” paths in the “training” step and $V_0^{N,K}$ is an estimate based on “exact” paths in both “training” and “testing” steps. Then

$$\begin{aligned}
\left| V_0 - \mathbb{E} \left[\tilde{V}_0^{N,K} \right] \right| &\leq \left| \mathbb{E} \left[\bar{V}_0^{N,K} \right] - \mathbb{E} \left[\tilde{V}_0^{N,K} \right] \right| \\
&\quad + \left| \mathbb{E} \left[\bar{V}_0^{N,K} \right] - \mathbb{E} \left[V_0^{N,K} \right] \right| + \left| \mathbb{E} \left[V_0 \right] - \mathbb{E} \left[V_0^{N,K} \right] \right| \\
&= R_1 + R_2 + R_3.
\end{aligned}$$

ESTIMATE FOR R_1 : First suppose for concreteness that $h_K > \gamma_K^{-1}$. We have

$$\begin{aligned} R_1 &= \left| \mathbb{E} \left[g_{\tau_K}(\tilde{Z}_{h_K, \tau_K}) - g_{\tilde{\tau}_K}(\tilde{Z}_{h_K, \tilde{\tau}_K}) \right] \right| \\ &\leq \mathbb{E} \sum_{l=0}^{\mathcal{J}-1} \left(1_{\{\tau_{K,l}=l, \tilde{\tau}_{K,l}>l\}} + 1_{\{\tau_{K,l}>l, \tilde{\tau}_{K,l}=l\}} \right), \end{aligned}$$

where

$$\tilde{\tau}_{K,l} = \inf\{l \leq j \leq \mathcal{J} : g_j(\tilde{Z}_{h_K,j}) > C_{K,j}(\tilde{Z}_{h_K,j})\}$$

and

$$\tau_{K,l} = \inf\{l \leq j \leq \mathcal{J} : g_j(Z_j) > C_{K,j}(Z_j)\}.$$

Introduce

$$\begin{aligned} \mathcal{E}_j &= \{g_j(Z_j) > C_{K,j}(Z_j), g_j(\tilde{Z}_{h_K,j}) \leq C_{K,j}(\tilde{Z}_{h_K,j})\} \\ &\cup \{g_j(Z_j) \leq C_{K,j}(Z_j), g_j(\tilde{Z}_{h_K,j}) > C_{K,j}(\tilde{Z}_{h_K,j})\}, \\ \mathcal{A}_{j,0} &= \left\{ 0 < |g_j(Z_j) - C_{K,j}(Z_j)| \leq \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\}, \\ \mathcal{A}_{j,i} &= \left\{ 2^{i-1} \kappa \sqrt{h_K} \log \frac{1}{h_K} < |g_j(Z_j) - C_{K,j}(Z_j)| \leq 2^i \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\} \end{aligned}$$

for $j = 0, \dots, \mathcal{J} - 1$, and $i > 0$. It holds

$$\begin{aligned} R_1 &\leq \mathbb{E} \left[\sum_{l=0}^{\mathcal{J}-1} 1_{\{\mathcal{E}_l\}} \right] \\ &= \mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{l=0}^{\mathcal{J}-1} 1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i}\}} \right]. \end{aligned}$$

Further, denote

$$\begin{aligned} \mathcal{D}_{l,i} &= \left\{ |g_l(Z_l) - C_{K,l}(Z_l)| \leq 2^i \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\}, \\ \mathcal{D}_{l,i}^* &= \left\{ |g_l(Z_l) - C_l^*(Z_l)| \leq 2^{i+1} \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\}, \\ \mathcal{B}_{l,i} &= \left\{ |C_{K,l}(\tilde{Z}_{h_K,l}) - C_{K,l}(Z_l)| > 2^{i-1} \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\}, \\ \mathcal{B}_{l,i}^* &= \left\{ |C_l^*(\tilde{Z}_{h_K,l}) - C_l^*(Z_l)| > 2^{i-2} \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\}, \\ \mathcal{S}_{K,l} &= \left\{ \sup_z |C_{K,l}(z) - C_l^*(z)| < 2^{i-3} \kappa \sqrt{h_K} \log \frac{1}{h_K} \right\} \in \mathcal{F}^K. \end{aligned}$$

Then it holds

$$\begin{aligned} \mathbb{E} \left[1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,0}\}} \right] &\leq \mathbb{P} \left(|g_l(Z_l) - C_l^*(Z_l)| \leq 2\kappa \sqrt{h_K} \log \frac{1}{h_K} \right) \\ &\quad + \mathbb{E} \left[\mathbb{P}^K \left(|C_{K,l}(Z_l) - C_l^*(Z_l)| \geq \kappa \sqrt{h_K} \log \frac{1}{h_K} \right) \right] \\ &\lesssim h_K^{\alpha/2} \log^\alpha \frac{1}{h_K} \end{aligned}$$

due to (MA). Analogously, using the fact that $|g_l(Z_l) - C_{K,l}(Z_l)| \leq |C_{K,l}(Z_l) - C_{K,l}(\tilde{Z}_{h_K,l})|$ on \mathcal{E}_l , we get

$$\begin{aligned} \mathbb{E} \left[1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i}\}} \right] &\leq \mathbb{E} \left[1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i} \cap \mathcal{S}_{l,i}\}} + 1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i} \cap \overline{\mathcal{S}}_{l,i}\}} \right] \\ &\leq \mathbb{E} \left[1_{\{\mathcal{B}_{l,i} \cap \mathcal{D}_{l,i} \cap \mathcal{S}_{l,i}\}} + 1_{\{\overline{\mathcal{S}}_{l,i}\}} \right] \\ &\leq \mathbb{E} \left[1_{\{\mathcal{B}_{l,i}^* \cap \mathcal{D}_{l,i}^*\}} + 1_{\{\overline{\mathcal{S}}_{l,i}\}} \right] \\ &= \mathbb{P} \left(\mathcal{B}_{l,i}^* \cap \mathcal{D}_{l,i}^* \right) + \mathbb{P}^K(\overline{\mathcal{S}}_{l,i}). \end{aligned}$$

Since

$$\mathbb{P}^K(\overline{\mathcal{S}}_{K,l,i}) \leq B_1 \exp \left(-2^{i-3} B_2 \kappa \log \frac{1}{h_K} \right) \leq B_1 h_K^{\alpha 2^{i-1}},$$

for $B_2 \kappa > 2\alpha$, we get

$$\mathbb{E} \left[\sum_{i=1}^{\infty} \sum_{l=0}^{\mathcal{J}-1} 1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i}\}} \right] \leq \sum_{i=1}^{\infty} \sum_{l=0}^{\mathcal{J}-1} \mathbb{P} \left(\mathcal{B}_{l,i}^* \cap \mathcal{D}_{l,i}^* \right) + O \left(h_K^{\alpha/2} \log^\alpha \frac{1}{h_K} \right).$$

Furthermore,

$$\begin{aligned} \mathbb{P} \left(\mathcal{B}_{l,i}^* \cap \mathcal{D}_{l,i}^* \right) &= \mathbb{P} \left(|\tilde{Z}_{h_K,l} - Z_l| > \mathcal{L}_C^{-1} 2^{i-2} \kappa \sqrt{h_K} \log \frac{1}{h_K}, |g_l(Z_l) - C_l^*(Z_l)| \leq 2^{i+1} \kappa \sqrt{h_K} \log \frac{1}{h_K} \right) \\ &= \mathbb{P} \left(|\tilde{Z}_{h_K,l} - Z_l| > \mathcal{L}_C^{-1} 2^{i-2} \kappa \sqrt{h_K} \log \frac{1}{h_K} \middle| \mathcal{D}_{l,i}^* \right) \mathbb{P} \left(\mathcal{D}_{l,i}^* \right), \end{aligned}$$

where

$$\mathbb{P} \left(\mathcal{D}_{l,i}^* \right) \leq 2^{\alpha(i+1)} \kappa^\alpha h_K^{\alpha/2} \log^\alpha \frac{1}{h_K}.$$

It remains to show that

$$\sum_{i=1}^{\infty} 2^{i\alpha} \sum_{l=0}^{\mathcal{J}-1} \mathbb{P} \left(|\tilde{Z}_{h_K,l} - Z_l| > 2^{i-2} \mathcal{L}_C^{-1} \sqrt{h_K} \log \frac{1}{h_K} \middle| \mathcal{D}_{l,i}^* \right) < \infty.$$

The Markov inequality implies for any $p > 0$

$$\mathbb{P} \left(\|\tilde{Z}_{h_K,l} - Z_l\| > 2^{i-2} \mathcal{L}_C^{-1} \sqrt{h_K} \log \frac{1}{h_K} \middle| \mathcal{D}_{l,i}^* \right) \leq \mathcal{L}_C^p 2^{-p(i-2)} \frac{\mathbb{E} \left[\sup_{l \leq j \leq \mathcal{J}} |\tilde{Z}_{h_K,j} - Z_j|^p \middle| \mathcal{D}_{l,i}^* \right]}{h_K^{p/2}}$$

and it is enough to prove that

$$\mathbb{E} \left[\sup_{l \leq j \leq \mathscr{J}} \left| \tilde{Z}_{h_K, j} - Z_j \right|^p \middle| \mathscr{D}_{l, i}^* \right] \leq C_p h_K^{p/2} \quad (6.1)$$

for any $p > 0$ and some constant $C_p > 0$. Under the condition

$$\mathbb{E} \left[\sup_{l \leq j \leq \mathscr{J}} |Z_j|^p \middle| \mathscr{D}_{l, i}^* \right] < \infty$$

the inequality (6.1) follows from the well known results on the strong convergence of discretisation schemes, see e.g., Kloeden and Platen (1992), Section 10.6.

ESTIMATE FOR R_2 : It holds

$$\begin{aligned} R_2 &= \left| \mathbb{E} \left[g_{\tau_K}(Z_{\tau_K}) - g_{\tau_K}(\tilde{Z}_{h_K, \tau_K}) \right] \right| \\ &\leq \mathscr{L}_g \left| \mathbb{E} \left[\sup_{j=0, \dots, \mathscr{J}} \|Z_j - \tilde{Z}_{h_K, j}\| \right] \right| \\ &\lesssim \sqrt{h_K}, \quad K \rightarrow \infty. \end{aligned}$$

ESTIMATE FOR R_3 : Taking into account that

$$C_l^*(Z_l) = \mathbb{E}^{\mathscr{F}_l} \left[g_{\tau_{l+1}^*}(Z_{\tau_{l+1}^*}) \right] < g_l(Z_l)$$

on $\{\tau_l^* = l\}$ and

$$C_l^*(Z_l) \geq g_l(Z_l)$$

on $\{\tau_l^* > l\}$, we get from Lemma 8

$$\begin{aligned} R_3 &= \left| \mathbb{E} \left[g_{\tau_K^*}(Z_{\tau_K^*}) - g_{\tau_K}(Z_{\tau_K}) \right] \right| \\ &\leq \mathbb{E} \left[\sum_{l=0}^{\mathscr{J}-1} |C_l^*(Z_l) - g_l(Z_l)| \left(1_{\{\tau_{K, l} = l, \tilde{\tau}_{K, l} > l\}} + 1_{\{\tau_{K, l} > l, \tilde{\tau}_{K, l} = l\}} \right) \right]. \end{aligned}$$

Introduce

$$\begin{aligned} \mathscr{E}_j &= \{g_j(Z_j) > C_{K, j}^*(Z_j), g_j(Z_j) \leq C_{K, j}(Z_j)\} \\ &\quad \cup \{g_j(Z_j) \leq C_K^*(Z_j), g_j(Z_j) > C_{K, j}(Z_j)\}, \\ \mathscr{A}_{j, 0} &= \left\{ 0 < \left| g_j(Z_j) - C_j^*(Z_j) \right| \leq \gamma_K^{-1/2} \right\}, \\ \mathscr{A}_{j, i} &= \left\{ 2^{i-1} \gamma_K^{-1/2} < \left| g_j(Z_j) - C_j^*(Z_j) \right| \leq 2^i \gamma_K^{-1/2} \right\} \end{aligned}$$

for $j = 0, \dots, \mathcal{J} - 1$ and $i > 0$. It holds

$$\begin{aligned}
R_3 &\leq \mathbb{E} \left[\sum_{l=0}^{\mathcal{J}-1} |C_l^*(Z_l) - g_l(Z_l)| 1_{\{\mathcal{E}_l\}} \right] \\
&= \mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{l=0}^{\mathcal{J}-1} |C_l^*(Z_l) - g_l(Z_l)| 1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i}\}} \right] \\
&= \gamma_K^{-1/2} \sum_{l=0}^{\mathcal{J}-1} \mathbb{P} \left(|g_l(Z_l) - C_l^*(Z_l)| \leq \gamma_K^{-1/2} \right) \\
&\quad + \mathbb{E} \left[\sum_{i=1}^{\infty} \sum_{l=0}^{\mathcal{J}-1} |C_l^*(Z_l) - g_l(Z_l)| 1_{\{\mathcal{E}_l \cap \mathcal{A}_{l,i}\}} \right].
\end{aligned}$$

Using the fact that $|g_l(Z_l) - C_l^*(Z_l)| \leq |C_l(Z_l) - C_l^*(Z_l)|$ on \mathcal{E}_l , we derive

$$\begin{aligned}
R_3 &\leq \gamma_K^{-1/2} \sum_{l=0}^{\mathcal{J}-1} \mathbb{P} \left(|g_l(Z_l) - C_l^*(Z_l)| \leq \gamma_K^{-1/2} \right) \\
&\quad + \sum_{i=1}^{\infty} 2^i \gamma_K^{-1/2} \mathbb{E} \left[\sum_{l=0}^{\mathcal{J}-1} 1_{\{|g_l(Z_l) - C_l^*(Z_l)| \leq 2^i \gamma_K^{-1/2}\}} \mathbb{P}^K \left(|C_{K,l}(Z_l) - C_l^*(Z_l)| > 2^{i-1} \gamma_K^{-1/2} \right) \right] \\
&\leq A \mathcal{J} \gamma_K^{-(1+\alpha)/2} + A \mathcal{J} \gamma_K^{-(1+\alpha)/2} \sum_{i=1}^{\infty} 2^i B_1 \exp(-B_2 2^{i-1}).
\end{aligned}$$

6.2 PROOF OF THEOREM 6

The proof follows the same lines as one of Theorem 4.

6.3 PROOF OF THEOREM 7

In order to simplify the notations, we use l instead of k_l . Also, we write $x \lesssim y$ if there exists a constant $c > 0$ that does not depend on $\beta, L, N_0, \dots, N_L$ such that $x \leq c \cdot y$. Moreover, $x \gtrsim y$ means $y \lesssim x$, and $x \asymp y$ stands for $x \lesssim y$ and $x \gtrsim y$. Let us consider the following optimization problem:

$$\sum_{l=0}^L k_l^x n_l h_l^{-1} \rightarrow \min_{\beta, L, N_0 \dots N_L} \quad (6.2)$$

with constraints

$$\begin{aligned}
&\gamma_L^{(1+\alpha)/2} + \left(\gamma_L \log^2 \frac{1}{\gamma_L} \vee h_L \log^2 \frac{1}{h_L} \right)^{\alpha/2} + h_L^{1/2} \lesssim \varepsilon \\
&\frac{1}{n_0} + \sum_{l=1}^L \left(\left(\gamma_{l-1} \log^2 \frac{1}{\gamma_{l-1}} \vee h_{l-1} \log^2 \frac{1}{h_{l-1}} \right)^{\alpha/2} + h_{l-1} \right) / n_l \lesssim \varepsilon^2.
\end{aligned}$$

We start with some simplifications. First of all, based on the special type of the functional (6.2) we will modify the constraints for the bias as

$$\max \left\{ \gamma_L^{(1+\alpha)/2}, \left(\gamma_L \log^2 \frac{1}{\gamma_L} \vee h_L \log^2 \frac{1}{h_L} \right)^{\alpha/2}, h_L^{1/2} \right\} \asymp \varepsilon$$

and for the variance

$$\max \left\{ \frac{1}{n_0}, \sum_{l=1}^L \frac{\left(\gamma_{l-1} \log^2 \frac{1}{\gamma_{l-1}} \vee h_{l-1} \log^2 \frac{1}{h_{l-1}} \right)^{\alpha/2}}{n_l}, \sum_{l=1}^L \frac{h_{l-1}}{n_l} \right\} \asymp \varepsilon^2.$$

which immediately implies $n_0 \asymp \varepsilon^{-2}$. We will always assume, that k_0 is sufficiently large, so that $\gamma_{l-1} \log^2 \frac{1}{\gamma_{l-1}}$ and $h_{l-1} \log^2 \frac{1}{h_{l-1}}$ are monotone functions with respect to l . Moreover, our analysis is carried out for sufficiently small ε . Now we can start an optimization procedure. We solve the problem in several steps.

STEP 1. $\beta \leq \mu$. We have $\beta \leq \mu \Rightarrow h_l \geq \gamma_l$ so the constraints can be rewritten as

$$\max \left\{ h_L^{\alpha/2} \log^\alpha \frac{1}{h_L}, h_L^{1/2} \right\} \asymp \varepsilon \Rightarrow \begin{cases} h_L^{\alpha/2} \log^\alpha \frac{1}{h_L} \lesssim h_L^{1/2} \asymp \varepsilon, & \text{if } \alpha > 1 \\ h_L^{1/2} \lesssim h_L^{\alpha/2} \log^\alpha \frac{1}{h_L} \asymp \varepsilon, & \text{if } \alpha \leq 1 \end{cases}$$

and

$$\begin{aligned} \max \left\{ \sum_{l=1}^L \frac{h_{l-1}^{\alpha/2} \log^\alpha \frac{1}{h_{l-1}}}{n_l}, \sum_{l=1}^L \frac{h_{l-1}}{n_l} \right\} &\asymp \varepsilon^2 \Rightarrow \\ &\Rightarrow n_l \asymp \varepsilon^{-2} L \max \left\{ h_{l-1}^{\alpha/2} \log^\alpha \frac{1}{h_{l-1}}, h_{l-1} \right\}. \end{aligned}$$

Now we will transform (6.2) via

$$\begin{aligned} \sum_{l=0}^L k_l^\chi n_l h_l^{-1} &\asymp \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ h_{l-1}^{\alpha/2} \log^\alpha \frac{1}{h_{l-1}}, h_{l-1} \right\} h_l^{-1} \asymp \\ &\asymp \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ h_l^{\alpha/2-1} \log^\alpha \frac{1}{h_l}, 1 \right\}, \end{aligned}$$

which leads us to the three cases:

1. $\alpha > 2$

$$\varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ h_l^{\alpha/2-1} \log^\alpha \frac{1}{h_l}, 1 \right\} \lesssim \varepsilon^{-2} L \kappa^{Lx} \asymp \varepsilon^{-2-\frac{2x}{\mu}} \log_\kappa \varepsilon^{-1},$$

$$\text{given } \beta = \mu \text{ and } L = \left\lceil \frac{2}{\mu} \log_\kappa \varepsilon^{-1} \right\rceil.$$

2. $1 < \alpha \leq 2$

$$\begin{aligned} \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ h_l^{\alpha/2-1} \log^\alpha \frac{1}{h_l}, 1 \right\} &\lesssim \varepsilon^{-2} L \kappa^{L(x+\beta\frac{2-\alpha}{2})} \log_\kappa^\alpha \varepsilon^{-2} \asymp \\ &\asymp \varepsilon^{-2-2\frac{2-\alpha}{2}} L \kappa^{Lx} \log_\kappa^\alpha \varepsilon^{-2} \asymp \varepsilon^{-4+\alpha-\frac{2x}{\mu}} \log_\kappa^{(\alpha+1)} \varepsilon^{-1} \end{aligned}$$

$$\text{given } \beta = \mu \text{ and } L = \left\lceil \frac{2}{\mu} \log_\kappa \varepsilon^{-1} \right\rceil.$$

3. $0 < \alpha \leq 1$

In this case, we start from analyzing the constraint for the bias error. It holds

$$h_L^{\alpha/2} \log^\alpha \frac{1}{h_L} \asymp \varepsilon \Rightarrow (L\beta)^{-2} \kappa^{L\beta} \asymp \varepsilon^{\frac{-2}{\alpha}}.$$

Based on the trivial fact, that $\log_\kappa y + (m+1)\log_\kappa (\log_\kappa y)$ is an upper bound for the solution $x = x(y)$ of the equation

$$\frac{\kappa^x}{x^m} = y, \quad \kappa > 1, y \gg 1, m \in \mathbb{N},$$

we have

$$\log_\kappa \varepsilon^{\frac{-2}{\alpha}} + 2 \log_\kappa \left(\log_\kappa \varepsilon^{\frac{-2}{\alpha}} \right) \lesssim L\beta \lesssim \log_\kappa \varepsilon^{\frac{-2}{\alpha}} + 3 \log_\kappa \left(\log_\kappa \varepsilon^{\frac{-2}{\alpha}} \right).$$

The latter inequality gives:

$$\begin{aligned} \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ h_l^{\alpha/2-1} \log^\alpha \frac{1}{h_l}, 1 \right\} &\lesssim \varepsilon^{-2} L \kappa^{L(x+\beta)} h_L^{\alpha/2} \log^\alpha \frac{1}{h_L} \\ &\asymp \varepsilon^{-1} L \kappa^{L(x+\beta)} \\ &\lesssim \varepsilon^{-1-\frac{2}{\alpha}} L \kappa^{Lx} \log_\kappa^3 \varepsilon^{-1} \\ &\lesssim \varepsilon^{-1-\frac{2}{\alpha}-\frac{2x}{\mu\alpha}} \log_\kappa^4 \varepsilon^{-1} \end{aligned}$$

$$\text{under the choice } \beta = \mu \text{ and } L = \left\lceil \frac{2}{\mu\alpha} \log_\kappa \varepsilon^{-1} + \frac{3}{\mu} \log_\kappa \left(\log_\kappa \varepsilon^{\frac{-2}{\alpha}} \right) \right\rceil.$$

STEP 2. $\beta > \mu$. We have $\beta \leq \mu \Rightarrow h_l \leq \gamma_l$ so the constraints can be rewritten as

$$\max \left\{ \gamma_L^{\alpha/2} \log^\alpha \frac{1}{\gamma_L}, h_L^{1/2} \right\} = \max \left\{ \kappa^{-L\mu\alpha/2} \log^\alpha \kappa^{L\mu}, \kappa^{-L\beta/2} \right\} \asymp \varepsilon$$

and

$$\begin{aligned} \max \left\{ \sum_{l=1}^L \frac{\gamma_{l-1}^{\alpha/2} \log^\alpha \frac{1}{\gamma_{l-1}}}{n_l}, \sum_{l=1}^L \frac{h_{l-1}}{n_l} \right\} &\asymp \varepsilon^2 \Rightarrow \\ &\Rightarrow n_l \asymp \varepsilon^{-2} L \max \left\{ \gamma_{l-1}^{\alpha/2} \log^\alpha \frac{1}{\gamma_{l-1}}, h_{l-1} \right\}. \end{aligned}$$

Now we will transform (6.2) in the same way, as we did before:

$$\begin{aligned} \sum_{l=0}^L k_l^x n_l h_l^{-1} &\asymp \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ \gamma_{l-1}^{\alpha/2} \log^\alpha \frac{1}{\gamma_{l-1}}, h_{l-1} \right\} h_l^{-1} \asymp \\ &\asymp \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ \kappa^{l(\beta-\mu\alpha/2)} \log^\alpha \frac{1}{\gamma_l}, 1 \right\}. \end{aligned}$$

It is clear, that $\kappa^{l(\beta-\mu\alpha/2)} \log^\alpha \frac{1}{\gamma_l} \gtrsim 1$, if $\alpha \leq 2$. Once again, we consider several cases

1. $0 < \alpha \leq 1 \Rightarrow \beta \geq \mu\alpha$. We have

$$h_L^{1/2} \lesssim \gamma_L^{\alpha/2} \log^\alpha \frac{1}{\gamma_L} \asymp \varepsilon \Rightarrow L \lesssim \frac{2}{\mu\alpha} \log_\kappa \varepsilon^{-1} + \frac{3}{\mu} \log_\kappa \left(\log_\kappa \varepsilon^{-\frac{2}{\alpha}} \right)$$

and

$$\begin{aligned} \sum_{l=0}^L k_l^x n_l h_l^{-1} &\lesssim \varepsilon^{-2} L \kappa^{L(x+\beta-\mu\alpha/2)} \log^\alpha \kappa^{\mu L} \asymp \\ &\asymp \varepsilon^{-1} L \kappa^{L(x+\beta)} \lesssim \varepsilon^{-1-2\frac{x+\mu}{\mu\alpha}} \log_\kappa^{1+\frac{3}{\mu}} \varepsilon^{-1}, \end{aligned}$$

$$\text{given } \beta = \mu \text{ and } L = \left\lceil \frac{2}{\mu\alpha} \log_\kappa \varepsilon^{-1} + \frac{3}{\mu} \log_\kappa \left(\log_\kappa \varepsilon^{-\frac{2}{\alpha}} \right) \right\rceil.$$

2. $1 < \alpha \leq 2$ and $\beta \geq \mu\alpha > \mu$

Just like in the previous case, we have

$$L \lesssim \frac{2}{\mu\alpha} \log_\kappa \varepsilon^{-1} + \frac{3}{\mu} \log_\kappa \left(\log_\kappa \varepsilon^{-\frac{2}{\alpha}} \right)$$

and hence

$$\begin{aligned} \sum_{l=0}^L k_l^\chi n_l h_l^{-1} &\lesssim \varepsilon^{-2} L \kappa^{L(\chi+\beta-\mu\alpha/2)} \log^\alpha \kappa^{\mu L} \asymp \\ &\asymp \varepsilon^{-1} L \kappa^{L(\chi+\beta)} \lesssim \varepsilon^{-3-\frac{2\chi}{\mu\alpha}} \log_\kappa^{1+\frac{3}{\mu}} \varepsilon^{-1}, \end{aligned}$$

given $\beta = \mu\alpha$ and $L = \left\lceil \frac{2}{\mu\alpha} \log_\kappa \varepsilon^{-1} + \frac{3}{\mu} \log_\kappa \left(\log_\kappa \varepsilon^{-\frac{2}{\alpha}} \right) \right\rceil$.

3. $1 < \alpha \leq 2$ and $\mu \leq \beta \leq \mu\alpha \Rightarrow 2\beta - \mu\alpha \geq 0$. Imposed constraints lead to the bias estimate:

$$\gamma_L^{\alpha/2} \log^\alpha \frac{1}{\gamma_L} \lesssim h_L^{1/2} \asymp \varepsilon \Rightarrow L\beta \asymp \log_\kappa \varepsilon^{-2}.$$

So for the total complexity we have

$$\begin{aligned} \sum_{l=0}^L k_l^\chi n_l h_l^{-1} &\lesssim \varepsilon^{-2} L \kappa^{L(\chi+\beta-\mu\alpha/2)} \asymp \varepsilon^{-4-\frac{2}{\beta}(\chi-\frac{\mu\alpha}{2})} \log_\kappa \varepsilon^{-1} \asymp \\ &\asymp \begin{cases} \varepsilon^{-3-\frac{2\chi}{\mu\alpha}} \log_\kappa \varepsilon^{-1}, & \text{if } \chi - \frac{\mu\alpha}{2} \geq 0 \text{ given } \beta = \mu\alpha \\ \varepsilon^{-4+\alpha-\frac{2\chi}{\mu}} \log_\kappa \varepsilon^{-1}, & \text{if } \chi - \frac{\mu\alpha}{2} \leq 0 \text{ given } \beta = \mu \end{cases} \end{aligned}$$

and $L = \left\lceil \frac{2}{\beta} \log_\kappa \varepsilon^{-1} \right\rceil$.

4. $\alpha > 2$ and $\beta \geq \mu\alpha$.

The answer will be the same, as in the case 2.

5. $\alpha > 2$ and $\mu\alpha \geq \beta \geq \mu$.

In the same way, as in the case 3 we have $L\beta \asymp \log_\kappa \varepsilon^{-2}$ and for the complexity estimate we consider two cases:

$$\begin{aligned} \sum_{l=0}^L k_l^\chi n_l h_l^{-1} &\asymp \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx} \max \left\{ \kappa^{l(\beta-\mu\alpha/2)} \log^\alpha \frac{1}{\gamma_l}, 1 \right\} \\ &\asymp \begin{cases} \varepsilon^{-2} L \sum_{l=0}^L \kappa^{lx}, & \text{if } \beta < \frac{\mu\alpha}{2} \\ \varepsilon^{-2} L \sum_{l=0}^L \kappa^{l(\chi+\beta-\mu\alpha/2)} \log^\alpha \frac{1}{\gamma_l}, & \text{else} \end{cases} \\ &\lesssim \begin{cases} \varepsilon^{-2} L \kappa^{L\chi}, & \text{if } \beta < \frac{\mu\alpha}{2} \\ \varepsilon^{-2} L \kappa^{L(\chi+\beta-\mu\alpha/2)} \log^\alpha \frac{1}{\gamma_L}, & \text{else} \end{cases} \\ &\lesssim \begin{cases} \varepsilon^{-2-\frac{4\chi}{\mu\alpha}} \log_\kappa^{1+\alpha} \varepsilon^{-1}, & \text{if } \frac{2\chi}{\mu\alpha} \leq 1, \text{ given } \beta = \frac{\mu\alpha}{2} \\ \varepsilon^{-3-\frac{2\chi}{\mu\alpha}} \log_\kappa^{1+\frac{3}{\mu}} \varepsilon^{-1}, & \text{if } \frac{2\chi}{\mu\alpha} \geq 1, \text{ given } \beta = \mu\alpha \end{cases} \end{aligned}$$

and $L = \left\lceil \frac{2}{\beta} \log_\kappa \varepsilon^{-1} \right\rceil$

After gathering the results from Step 1 and 2 we obtain the complexity 4.

REFERENCES

- L. Andersen: A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor Market Model. *J. Computat. Financ.*, **3**, 5–32 (2000).
- D. Belomestny: Pricing Bermudan options using nonparametric regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, **15**(4), 655–683 (2011).
- D. Belomestny, J. Schoenmakers and F. Dickmann: Multilevel dual approach for pricing American style derivatives, to appear in *Finance and Stochastics* (2013).
- M. Broadie and P. Glasserman: Pricing American-style securities using simulation. *J. Econ. Dyn. Con.*, **21**, 1323–1352 (1997).
- M. Broadie, P. Glasserman: A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, **7**(4), 35–72 (2004).
- J. Carriere: Valuation of early-exercise price of options using simulations and nonparametric regression. *Insur. Math. Econ.*, **19**, 19–30 (1996).
- D. Egloff: Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, **15**, 1396–1432 (2005).
- M.B. Giles: Multilevel Monte Carlo path simulation. *Operations Research* **56**(3), 607–617 (2008).
- E. Giné and A. Guillou: A law of the iterated logarithm for kernel density estimators in the presence of censoring. *Ann. I. H. Poincaré*, **37**, 503–522 (2001).
- P. Glasserman: Monte Carlo Methods in Financial Engineering. In: Springer (2004).
- P. Glasserman and B. Yu: Number of Paths Versus Number of Basis Functions in American Option Pricing. *Ann. Appl. Probab.*, **14**, 2090–2119 (2004).
- P. Kloeden and E. Platen: Numerical solution of stochastic differential equations. Applications of Mathematics (New York), 23. Springer-Verlag, Berlin (1992).
- F. Longstaff and E. Schwartz: Valuing American options by simulation: a simple least-squares approach. *Rev. Financ. Stud.*, **14**, 113–147 (2001).
- J. Tsitsiklis and B. Van Roy: Regression methods for pricing complex American style options. *IEEE Trans. Neural. Net.*, **12**, 694–703 (1999).